

1-1-2013

Developing an Understanding of Data Management Education: A Report from the Data Information Literacy Project

Jake Carlson

Purdue University, jakecarlson@purdue.edu

Lisa Johnston

University of Minnesota - Twin Cities, ljohnsto@umn.edu

Brian Westra

University of Oregon, bwestra@uoregon.edu

Mason Nichols

Purdue University, masonnichols@purdue.edu

Follow this and additional works at: http://docs.lib.purdue.edu/lib_fspres



Part of the [Curriculum and Instruction Commons](#), [Educational Assessment, Evaluation, and Research Commons](#), [Library and Information Science Commons](#), and the [Other Education Commons](#)

Recommended Citation

Carlson, Jake; Johnston, Lisa; Westra, Brian; and Nichols, Mason, "Developing an Understanding of Data Management Education: A Report from the Data Information Literacy Project" (2013). *Libraries Faculty and Staff Presentations*. Paper 11.
http://docs.lib.purdue.edu/lib_fspres/11

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

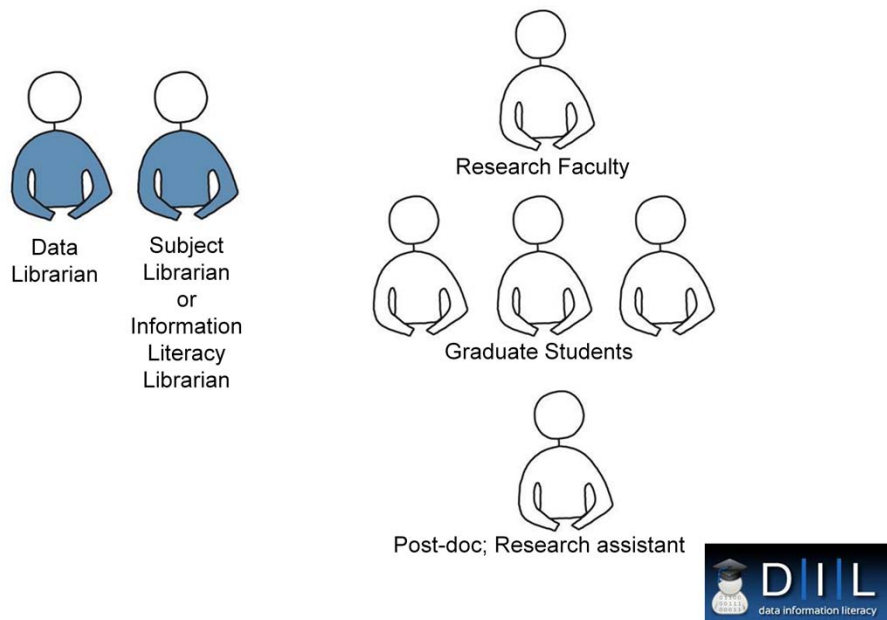
Developing an Understanding of Data
Management Education: A Report from the
**Data Information Literacy
Project**

Jake Carlson, Lisa Johnston,
Brian Westra, Mason Nichols

<http://www.datainfoilit.org>



Project Structure








Structure of the project

- Data librarian, subject librarian, faculty group

and Goals:

- Identify DIL skills appropriate to disciplinary contexts,
- Build infrastructure and capacity for teaching DIL skills,
- Develop a robust model for librarians to articulate DIL curricula in their research communities.

Five Case Studies

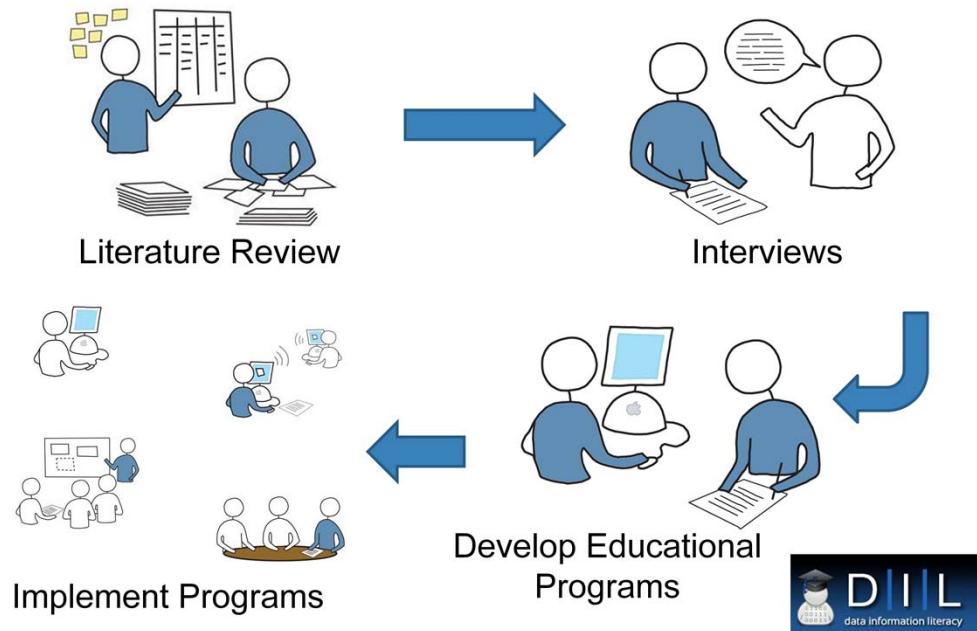
				
Cornell	Minnesota	Oregon	Purdue #1	Purdue #2
Natural Resources	Civil Engineering	Ecology	Electrical & Computer Engineering	Agricultural & Biological Engineering
Longitudinal data of fisheries and water quality	Real-time sensor data on bridge structures	Climate change and plant growth data	Software code in community service projects	Simulation data of hydrological processes

Try to add animation so that each column appears one at a time?
 IL=Information Literacy Librarian?

P1: Service Learning Center
 O: Grant Funded Project
 M: Research Lab

Field Work with Plants and Climate
 Bridge Construction/ Real-time bridge sensors
 C. Longitudinal Fisheries

Project Phases



The DIL project has two phases:

Phase 1: Disciplinary Case Studies

- Literature Reviews / Environmental Scans
 - How does the literature in the discipline address data management / curation issues?
- Interviews
 - Using a modified version of the Data Curation Profile to uncover current practices and perceived needs.
 - Faculty Member
 - Graduate Students
- Develop/Implement Educational Programming

Phase 2. Developing a Model for DIL

- Symposium

Literature Review



Understanding Disciplinary...

- Concepts of Data and Data Management Issues
- Terminology
- Best Practices / Standards
- Educational Approaches



NEW SECTION

We began by conducting literature reviews in the disciplines of our faculty partners to uncover how the 12 competencies were described and addressed. We shared information via Mendeley. One important outcome of the literature reviews was a recognition of the need to clarify our definitions of the 12 competencies and our approach in using these competencies as touchstones in the subsequent interviews. The faculty and students participating in this project may have different understandings and definitions of the competencies based on their experiences and background. Therefore, rather than assigning strict definitions, we described each by listing activities and skills that would reflect the nature of the competency.

We didn't try to do this on our own. We wanted to align our efforts with the fields' vision of what data management best practices were. For example, for software code they don't refer to provenance in terms of tracking the changes, they refer to tracking.

Interviews



Understanding Local:

- Data / Research
 - Lab Practices
 - Priorities
-
- Faculty (n = 8),
 - Grad Students (n = 15),
 - Research Assistants (n = 2)



Each of the project teams then conducted interviews. Eight of the interviews were with faculty. The other 17 interviews were with current or former graduate students of the interviewed faculty, or in two cases with a post-doc and a research assistant. The interviews were conducted in the spring and summer of 2012. The interview protocol was based on the structure of the Data Curation Profiles Toolkit developed at Purdue. The protocol consisted of an interview worksheet, which contained a series of questions for the interviewee to complete in writing during the interview, and an interviewer's manual, which contained follow up questions for the interviewer to ask based on the responses written by the interviewee. Our interview protocol is available for download from the project website.

We asked the interviewees to describe the data they create, describe their data life-cycle, react to the 12 competencies (prioritize)

Competencies of DIL

Processing and Analysis	Curation and Re-Use
Management and Organization	Conversion and Interoperability
Preservation	Visualization and Representation
Databases and Formats	Discovery and Acquisition
Ethics and Attribution	Metadata and Description
Quality and Documentation	Cultures of Practice

Carlson, J., Fosmire, M., Miller, C., & Nelson, M. S. (2011). Determining data information literacy needs: A study of students and research faculty. *portal: Libraries and the Academy*, 11, 629-657.
[doi:10.1353/pla.2011.0022](https://doi.org/10.1353/pla.2011.0022)



Numerous studies call for DIL needs

Interview Methodology

Module 7 – Organization and Description of Data

Please indicate how important you believe it is for your students to be knowledgeable in each of the competencies listed below by the time they graduate by circling a response below:

Interview Worksheet

Data Management and Organization

Skills may include:

Understands the lifecycle of data, develops data management plans, and keeps track of the relation of subsets or processed data to the original data sets. Creates standard operating procedures for data management and documentation.

1 2 3 4 5
Not Important Somewhat Important Important Very Important Essential
I don't know or NA



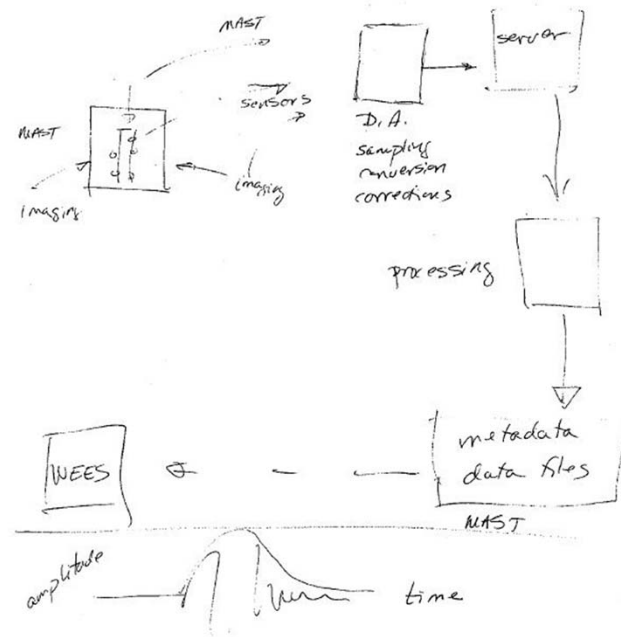
Interviewer's Manual



The DIL competency rankings show that, on average, participants valued each skill, as all of them were ranked as “Important” or higher. However, there was considerable variance in the responses received as indicated by the high standard deviations (ranging from .75 to 1.02).

As demonstrated in [Figure 2](#), there are noticeable differences in how the faculty and students participating in this project viewed some of the competencies. Overall, faculty placed a higher value on students developing competencies in actively working with data (“Data Processing and Analysis,” “Data Visualization and Representation”) and in competencies that would sustain the value of the data over time (“Metadata and Data Description,” “Data Quality and Documentation”) than the students did. Students indicated in the interviews that competencies in “Discovery and Acquisition” were important to them in learning their field and contextualizing their research. Two of the faculty, both of whom were working with code as their data, gave “Data Management and Organization” a lower ranking than the other participating faculty. One faculty believed that, individually, students should know how to manage their own data, but did not necessarily need to know how to develop systems or plans for larger units. The other found it difficult to respond, not knowing what constituted good management practice and if it would be worth investing in.

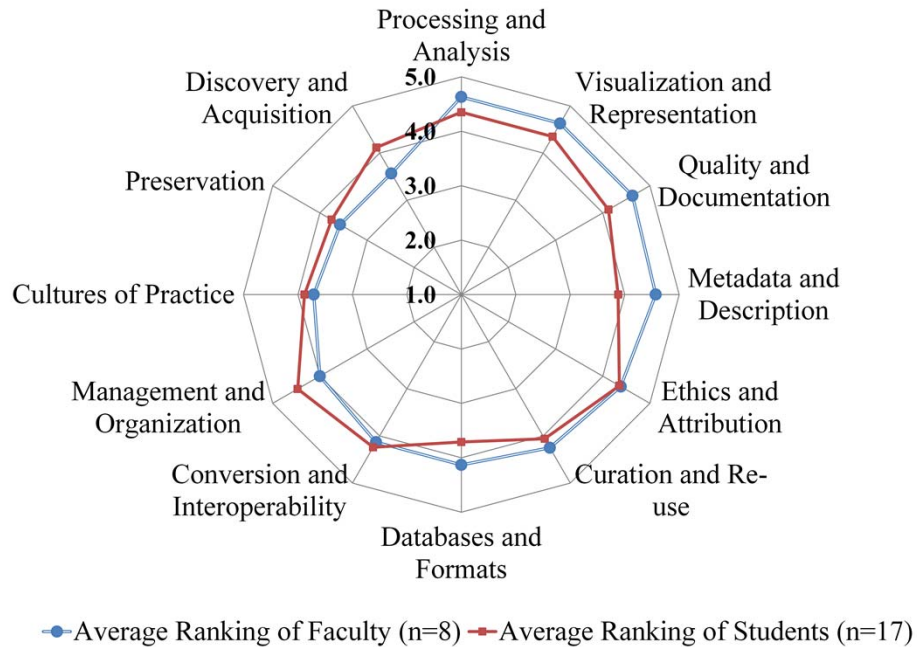
Interview Results



Each of the project teams then conducted interviews. Eight of the interviews were with faculty. The other 17 interviews were with current or former graduate students of the interviewed faculty, or in two cases with a post-doc and a research assistant. The interviews were conducted in the spring and summer of 2012. The interview protocol was based on the structure of the Data Curation Profiles Toolkit developed at Purdue. The protocol consisted of an interview worksheet, which contained a series of questions for the interviewee to complete in writing during the interview, and an interviewer's manual, which contained follow up questions for the interviewer to ask based on the responses written by the interviewee. Our interview protocol is available for download from the project website.

We asked the interviewees to describe the data they create, describe their data life-cycle, react to the 12 competencies (prioritize)

Rankings of Importance



The DIL competency rankings show that, on average, participants valued each skill, as all of them were ranked as “Important” or higher. However, there was considerable variance in the responses received as indicated by the high standard deviations (ranging from .75 to 1.02).

As demonstrated in [Figure 2](#), there are noticeable differences in how the faculty and students participating in this project viewed some of the competencies. Overall, faculty placed a higher value on students developing competencies in actively working with data (“Data Processing and Analysis,” “Data Visualization and Representation”) and in competencies that would sustain the value of the data over time (“Metadata and Data Description,” “Data Quality and Documentation”) than the students did. Students indicated in the interviews that competencies in “Discovery and Acquisition” were important to them in learning their field and contextualizing their research. Two of the faculty, both of whom were working with code as their data, gave “Data Management and Organization” a lower ranking than the other participating faculty. One faculty believed that, individually, students should know how to manage their own data, but did not necessarily need to know how to develop systems or plans for larger units. The other found it difficult to respond, not knowing what constituted good management practice and if it would be worth investing in.






Synthesis (Commonalities)

- Lack of formal training in data management
- Lack of formal policies in the research team
- Self-directed learning through trial and error
- Focus on data mechanics over deeper concepts



Analyzing the interview transcripts revealed several high-level commonalities across the five case studies. Among them, the overall lack of formal training, the lack of formal policies governing the data in the lab (team), self-directed learning through trial and error, and a general focus on mechanics over concepts.

Local Themes

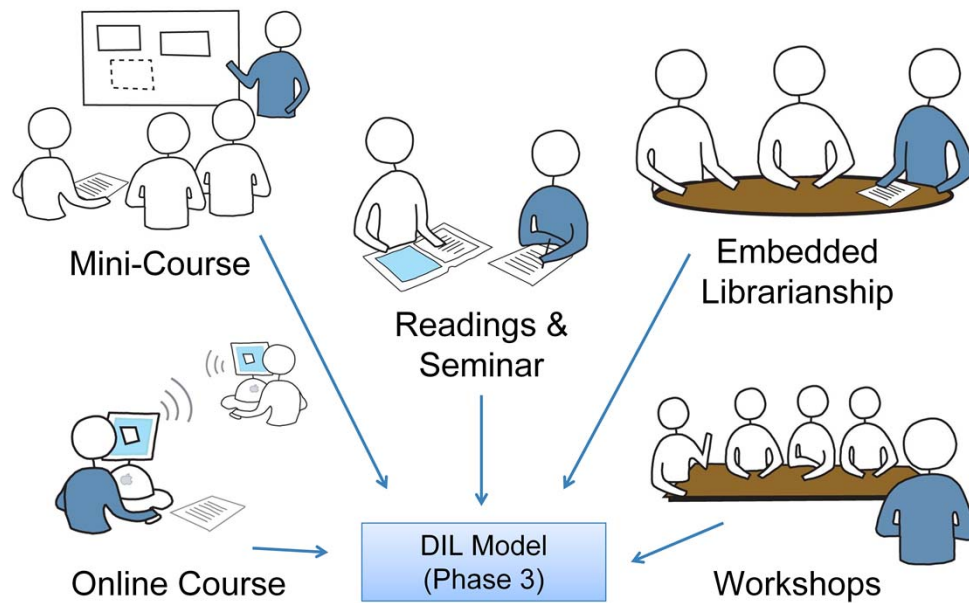
				
Cornell	Minnesota	Oregon	Purdue #1	Purdue #2
Natural Resources	Civil Engineering	Ecology	Electrical & Computer Engineering	Agricultural & Biological Engineering
Data sharing Databases Stewardship	Data ownership Long-term access	Cultures of practice Metadata Closing out a grant	Documentation & organization Transfer of responsibility	Standard operating procedures Metadata

Try to add animation so that each column appears one at a time?
IL=Information Literacy Librarian?

P1: Service Learning Center
O: Grant Funded Project
M: Research Lab

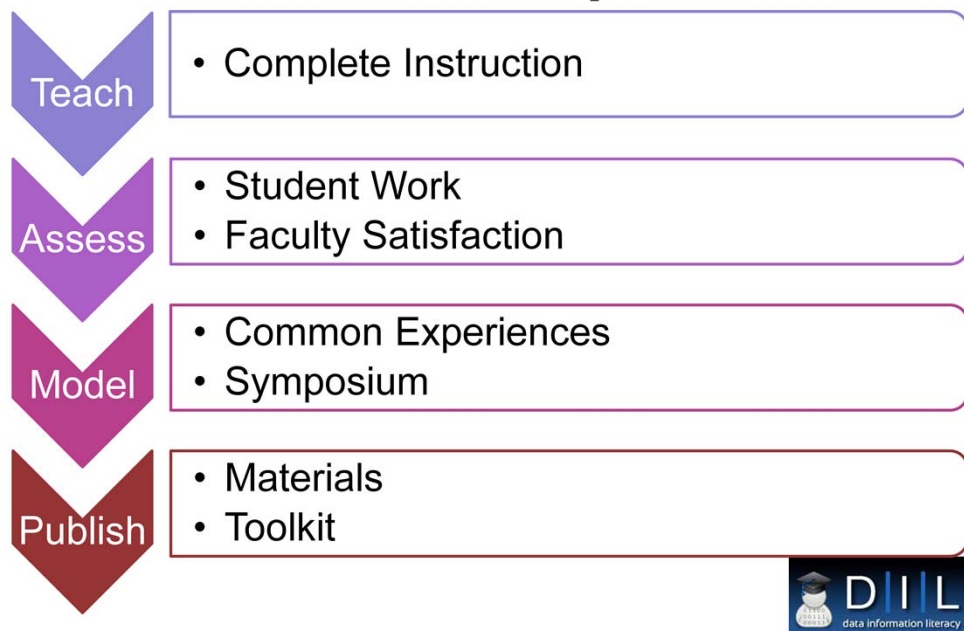
Field Work with Plants and Climate
Bridge Construction/ Real-time bridge sensors
C. Longitudinal Fisheries

Instructional Approaches



The DIL project resulted in 5 different approaches. These will help build our model for DIL.

Next Steps



Complete instruction at each institution

evaluation - ongoing throughout instruction plus some follow-up post instruction

draft model - early 2013

symposium - spring 2013

publish and disseminate model and toolkit - summer 2013

Credits

Principal Investigator:

- Jake Carlson - Purdue University



Co-Principal Investigators:

- Camille Andrews – Cornell University
- Marianne Stowell Bracke – Purdue University
- Michael Fosmire – Purdue University
- Jon Jeffryes – University of Minnesota
- Lisa Johnston – University of Minnesota
- Megan Sapp Nelson – Purdue University
- Dean Walton – University of Oregon
- Brian Westra – University of Oregon
- Sarah Wright – Cornell University

Grad Asst: Mason Nichols – Purdue University

The DIL project is supported by a National Leadership Grant from the Institute of Museum and Library Services (IMLS), LG-07-11-0232-11. Project personnel are from Purdue University: Marianne Stowell Bracke, Jake Carlson (PI), Michael Fosmire, Mason Nichols (Grad. Asst.) and Megan Sapp-Nelson; Cornell University: Camille Andrews and Sarah Wright; the University of Minnesota: Jon Jeffryes and Lisa Johnston; and the University of Oregon: Dean Walton and Brian Westra.

Thanks and Stay Tuned!

Jake Carlson - jrcarlso@purdue.edu

Lisa Johnston - ljohnsto@umn.edu

Brian Westra - bwestra@uoregon.edu

Mason Nichols - masonnichols@purdue.edu



datainfolit.org



[@datainfolit](https://twitter.com/datainfolit)



facebook.com/datainfolit